

## A Psychologist's View of Validating Aviation Systems

Earl S. Stein & Dan Wagner

United States Federal Aviation Administration

### Introduction

All systems, no matter what they are designed to do, have shortcomings that may make them less productive than was hoped during the initial development. Such shortcomings can arise at any stage of development: from conception to the end of the implementation life cycle. While systems failure and errors of a lesser magnitude can occur as a function of mechanical or software breakdown, the majority of such problems in aviation are usually laid on the shoulders of the human operator and, to a lesser extent, on human factors (Nagel, 1988). The operator bears the responsibility and blame even though, from a human factors perspective, error may have been designed into the system.

Human factors is not a new concept in aviation. The name may be new, but the issues related to operators in the loop date back to the industrial revolution of the nineteenth century and certainly to the aviation build-up for World War I. During this first global confrontation, military services from all sides discovered rather quickly that poor selection and training led to drastically increased personnel losses. While hardware design became an issue later, the early efforts were primarily focused on increased care in pilot selection and on their training. This actually involved early labor-intensive simulation, using such devices as sticks and chairs mounted on rope networks which could be manually moved in response to control inputs.

The use of selection criteria and improved training led to more viable person-machine systems. More pilots survived training and their first ten missions in the air, a rule of thumb arrived at by experience which predicted ultimate survival better than any other. This rule was to hold though World War II. At that time, personnel selection and training became very sophisticated based on previous standards. Also, many psychologists were drafted into Army Air Corps programs which were geared towards refining the human factor. However, despite the talent involved in these programs and the tremendous build-up of aviation during the war, there were still aircraft designs that were man killers (no sexism implied since all combat pilots were men). One classic design error that was identified fifty years ago was the multipointer altimeter, which could easily be misread especially by a pilot under considerable task load. It has led to flying fully operational aircraft into the terrain (Fitts and Jones, 1947). The authors of the research which formally identified this problem put "Human Errors" in quotes to express *their* dissatisfaction with the traditional approach to accident investigation. It traditionally places the burden of guilt on the operator. Some of these altimeters still exist in older aircraft to this day.

Human Factors Certification of Advanced Aviation Technologies  
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland  
Copyright © 1994 Embry-Riddle Aeronautical University Press

## Human Factors in Complex Systems

The airspace system has become increasingly more complicated since the Second World War, and an emphasis on aircraft issues alone would not do it service. The potential for chains of events leading to system breakdown has increased with the volume of traffic and the complexity of both air and ground subsystems. While the concept of human factors is not new, it is continually being rediscovered or ignored by systems developers. In addition to cockpit crew operations, the modern civil and military airspace system must include airport ground operations, conducted by air traffic control, and maintenance of both air and ground hardware/software. This latter area is handled by the airway facilities personnel in the FAA, while the former is a function of airlines, the FAA, fixed base operators, and other airframe and power maintenance resources.

All of these systems and subsystems have people working in them, and in cases like the multipointer altimeter, the hardware and software are not user friendly. Systems have often been created in which the operator was the last to know what was going on, and human factors professionals only became involved by exception when the designers knew they had a problem. In the latter case, it is not unusual for the designers to come to human factors people with a request to find a better way to select and train because the hardware and software designs are already frozen. This is a step back in time where human factors by whatever name was only viewed as useful from the limited perspectives of training and selection.

Warm and Dember (1986) described their concerns over systems that are designed in such a way that there may be attentional lapses to the point that operators are no longer "awake at the switch." In such a situation, whether in aviation or not, the system is operating in "free flight" without any supervisory control by human hand. The fact that this can and does happen became clear several years ago when the Nuclear Regulatory Commission closed the Peach Bottom power plant in Pennsylvania. Inspectors found operators literally asleep at their stations. Obviously this problem was not identified and addressed during the development of the plant.

An example of a situation in which systems designers decided to evaluate their product before it was too late was reported by Kantowitz and Sorkin (1983). A consumer electronics company was designing a new answering machine that would have been the answer to all needs of business and industry. The designers were convinced that they had a very marketable product which would be quite profitable. However, someone decided that it would be reasonable to test the product using a sample of people who were the intended users: secretaries. They gathered a group of ten and provided the documentation and the equipment and told them to go forth and use it. Not one could figure out how to operate the system. The designers concluded that the problem had to be training, so they rewrote the documentation and developed a basic training program. Another group of ten secretaries was mobilized. Of these ten only two could learn to operate the system. They both had previous backgrounds in computer programming. The designers had created a system for themselves and not for the users. Fortunately, they had chosen to test. Had they attempted to market the product it would have been a financial disaster.

Are there parallels between the answering machine example and the evolution of aviation systems? Of course! Thomas (1985) described the development of air traffic control over a nine year period in the 1950's and 60's. He noted the difficulty in transitioning from the older broad band radars to the more modern narrow band digital systems. Part of the problem was not so much a matter of system design but rather a function of preparation of the users to accept the

new equipment. Further, there were capabilities in the newer equipment that the operators tended to use or not based on their preferences and experience. At the time, very little consideration of the users was apparent in system development. This is changing slowly, and modern ATC systems often have controller panels involved in the design effort.

## The Concept of Certification

Certification is a legalistic term. It implies an organizational standing such that the certifier or certifying agency has the power to determine if a system can be used and under what conditions. Certification implies sound methods that have met the tests of time, validity, and reliability. It suggests protection of the public from hazards generated by systems that have been poorly designed and/or from operators that are unqualified or unable. Certification is a novel concept in human factors, at least from the viewpoint of some stand alone process separate and distinct from the engineering aspects of system development. Human factors professionals must ask themselves if they really want to become involved in certification and if so why?

In the United States, the Federal Aviation Administration already carries the responsibility for certifying aircraft and personnel in aviation. It also certifies its own equipment and the people who control and maintain the airspace system. Most of this legal requirement is handled by subject matter experts who are flight examiners, controllers, and hardware systems specialists. In the past, the role of human factors personnel has always been in the background and has been principally advisory to those who would actually sign off on an aircraft or other system. While human factors as a science may have been viewed as an information source, it has been the exception rather than rule for anyone in the business of certification to ask the advice and council of human factors specialists.

One partial exception is recalled by these authors. This was an effort to provide the designers of the Automated En Route Radar Air Traffic Control (AERA) with empirical validation of a construct called "workload probe." The probe was a computer algorithm that would theoretically predict controller workload up to 20 minutes in advance based on weather and anticipated traffic. It was tested in simulation at the FAA Technical Center. Measures of real time controller workload were collected using a Cooper-Harper type scale called the air traffic workload input technique (ATWIT). The results of the workload predictions from the probe were significantly correlated with participants self ratings using ATWIT and with ratings by over-the-shoulder observers. This proved that the concept of workload probe was viable, but it hardly qualified as certification since AERA as a system is not yet ready for certification testing. However, the process of empirical testing could be viewed as model for future systems evaluation. Human factors are here to stay in one form or another and has the support of Congress based on law.

The Aviation Safety Research Act of 1988 led to an increased awareness of the possibilities of human factors in the air space system. It required the FAA in particular to expend a finite portion of its annual budget in human factors related to new systems under development. One of the most visible products resulting from the Act has been the *National Plan for Human Factors* (FAA, 1991). This is a very comprehensive document which theoretically defines the human factors research needs for the present and the foreseeable future. One is struck by the magnitude of the document and the implication/admission that there is a great deal which is not

currently in the corporate body of knowledge concerning person-machine systems and subsystems in aviation.

Within the domain of aircraft certification, the plan states the following: "The FAA is responsible for the human factors evaluation and certification of aircraft. The personnel most responsible for this job are FAA certification pilots. These pilots are finding it very difficult to keep up with the human factors implications of the latest developments in flight deck automation and advanced technology aircraft." The suggested solution is the development of handbooks, checklists, and special courses.

While "certification" is cited in the chapter on aircraft from an aircraft perspective, conceptually it does not appear within the aircraft maintenance section of the plan. There are, however, a variety of research related issues to include the old standbys of selection and training as well as an expressed need to define the task structure more thoroughly.

The plan also notes that "... airway facilities personnel problems and needs have remained largely ignored" (FAA, 1991). These men and women work unseen behind the more glamorous positions of pilots and controllers. They maintain equipment, some of which retains vacuum tubes and most of which was designed without any reference to even common standards of ergonomic considerations. There are FAA standards of system maintainability, but it is notable that the plan does not address any issues related to certification of those standards or for that matter to the test equipment that personnel will use to accomplish the maintenance task. It does imply, however, that additional work will be necessary to evaluate such factors as maintenance documentation (an often ignored area of system operations) and the approaches to diagnostic support requirements.

It should be noted that airway facilities currently use a certification process. Before a new piece of equipment or a piece of equipment that was taken out of service for maintenance or repair is placed in service in the national airspace, it must be certified. That is, a piece of equipment, such as a radar, must be "certified" to be performing its intended function acceptably and within specified tolerances before it can be placed in service in the national airspace system. Of course, this does not mean the equipment is well designed, user-friendly, or even maintainable. It simply means that the required function is being accomplished.

The issues this process raises from an airways facilities perspective is that if a piece of equipment, a facility, or a system is accomplishing its intended function, how is human factors certification going to enhance this process? What are the criteria to be applied that say, "This piece of equipment or system is working better now (i.e., safer) than it would have if it were not human factors certified." Even if we are invited to certify newer items/systems designed to meet acceptable criteria, how can we certify older pieces of equipment that were not designed to human factors principles or standards?

These practical issues suggest that if certification is to become a reality, we should be prepared to support it for a long time, since existing equipment may still be in the field 10-20 years from now; and that a certification process must provide a value-added dimension (e.g. safety, reliability, maintainability, etc.), or else why do it at all. Additionally, certification of the individual cannot mean that the person has expertise in all areas of human factors, since the field has simply become too diverse to maintain proficiency in all domains.

Certification of air traffic control systems is traditionally accomplished by air traffic controllers and operational test and evaluation personnel. In the past, human factors personnel have been involved in research and development of new systems and to some extent in OT & E. However, human factors has not been considered a major element of the testing process but rather as a necessary adjunct or from a program managers perspective, a less than necessary evil. This view has been justified in the eyes of the system developers because when systems

were developed in the past without adequate human factors support, any analysis of prototypes was bound to identify unforeseen person-machine issues late in the development cycle. The human factors plan discusses many areas of needed research for evolving systems, but addresses certification primarily from the perspective of controller personnel issues.

Introducing human factors into the certification process beyond what is cited in aircraft systems already will require a cultural change of the magnitude invoked by the Department of the Army MANPRINT program (Booher, 1990). In the introduction of his book, Harold Booher writes:

People are both the cause and the solution. People are both the benefactors and the victims. Through human error in design, operation, or repair of machines, others are hurt killed or made unhappy or, at the least, inconvenienced. On the other hand, it is through human intelligence and unique human skills that equipment, organizations and knowledge enhancing products are designed and operated effectively, efficiently, and safely. (P. 2)

Booher sees the solution to these issues as a reorientation from hardware to people and conceives the mechanism to achieve this organizational change as "Total Quality Management."

In MANPRINT, the goal is to integrate human factors into every level of material development. This requires a complete systems view of each new piece of technology or hardware. It is mandated through Army regulations. This is both a program and a philosophy of system development. While the program is formally limited to the U.S. Army, the philosophy could go well beyond to other high reliability organizations in which small errors can lead to big problems. The MANPRINT philosophy suggests that there is a long term payback for good human factors in the initial development of a system. This is a life cycle approach to new technology. One of the problems identified by Booher (1990) is that the benefits of early investment in good design for long term system reliability do not usually accrue to the program managers personally because they move on to other developmental efforts. The key to success may be in educating all the personnel involved the process. This sounds reasonable in principle but is not easy to implement in practice. While there have been discussions within the Federal Aviation Administration of developing something similar to Army Regulation 602-2 (U.S. Army, 1990), to date there is no such document in place.

## **Human Factors Certification of Aviation Systems**

There are two central questions that are recurrent. First, do we want to be involved in the certification process to an extent beyond what exists now – which is principally an advisory role – when asked? Second, if we were able to see that increased participation was desired, could we live up to the challenge? Do we have the methods and measures to go beyond our episodic advisory role?

As stated earlier, aircraft systems certification is well institutionalized and has a history going back before World War II. However, as indicated by the National Plan, both engineering and human factors are theoretically accomplished by flight test pilots. This places a great deal of weight on their shoulders which they are very willing to bear. One advantage they have over other aviation systems, such as air traffic control and airway facilities, is that because of their

long history there are fairly well recognized standards of systems and individual performance in the cockpit. Aircraft systems are usually, although not always, designed with clear goals for what they are intended to do; this makes evaluation easier. Despite this, it will not be easy to increase the role of human factors in the overall certification process because this would involve a change in thinking on the part of the personnel currently doing the work. It could be viewed as an additional impediment to the process.

In ground systems, the situation is probably even more complicated. First, not all systems are designed with integrated goals other than the global desire to improve safety and performance while reducing operator workload. These are admirable yet non-specific goals and lead to the complexity of defining both system and individual performance. Can we define performance and use it as criterion for evaluation during development and operational testing? Past experience indicates that this is often a moving target which seems to progress as the system evolves. It is further complicated by moving beyond a general definition of performance with a qualitative description of what constitutes "good performance." This involves subjective decisions by subject matter experts who may well have differences of opinion. One can only evaluate the impact of a new system if it can be determined that it somehow is worth the investment in time, money, and safety. This improvement can only be measured if there is technical agreement on what "improvement" means.

From a human factors standpoint, it is not adequate to conclude that a system simply reduces workload because it may in fact reduce the load to the detriment of situational awareness or general alertness. It is the system performance that is the key, and the human is a critical component of the system that can and should not be equated with a piece of hardware. There are issues which machines do not become involved in such as motivation, esprit de corps, fatigue, and human information processing. All of these factors can have an impact on both human and systems performance.

In the process of evaluating a system, there are very few standards which apply consistently. Simply meeting the minimum requirements under MIL-STD 1472D (DOD, 1981) may not be nearly enough to adequately certify a specific application. For example, 1472D is purposefully vague when it comes to human workload. It says that we should not overload the operator. What constitutes an overload would vary from one application to another and for that matter from one operator to another. There have been discussions concerning rewriting this document for civilian aviation applications but no document has been circulated yet, although an FAA airways facilities human factors design standard has been written and is presently under internal review. Even when it is completed, it is likely that systems evaluation and subsequent certification will have to be accomplished on a case by case basis focusing on the design goals of each. For ground systems in particular, this will mean that considerable effort will be required in order to identify and validate suitable metrics that not only meet the criteria necessary for good measurement but have obvious relationships to performance of the systems being evaluated.

Falling back on human factors handbooks and data from days past is useful during design and early development of systems, but these authors believe that the bottom line of any system should be how it really performs either in actual prototype or preferably in high fidelity simulation before prototype testing is begun in the field.

To do either type of testing requires empirical and high quality measurement, and to date there is no general agreement on either the measures themselves or what constitutes acceptable performance. This is especially true when a system is under development to replace one which is currently operating. It can be anticipated that system change may be finite but not meteoric in magnitude. The more subtle the anticipated differences, the more difficult it will be to

demonstrate them using conventional measurement and statistical tools. While empirical testing offers the evaluators an opportunity to reach out for what could be a better estimate of operational reality, it does have its drawbacks: time and cost. It takes longer and it costs more to follow a philosophy like MANPRINT in which human factors is integrated into the developmental cycle and empirical evaluation is accomplished whenever possible (at least in the ideal model of the philosophy). What this all means is that considerable effort must be expended to develop the measurement tools that would provide adequate credibility for human factors in the certification process. While there is a history in air side human factors, there is a very limited but fortunately growing level of expertise in ground side operations.

### **Criteria for Human Factors in Certification**

Once again the questions are: Do we in the civil aviation human factors community want to become involved in the certification process and are we able to produce?

The answer the first question is: Of course we do! Even if we do not have all the answers today and all the tools of tomorrow, we can still be of more help than we were in the past if we are invited into the development cycle sooner. The answer to the second question is a qualified maybe. We need to develop ground side and improve air side measurement. This will not happen overnight despite the belief by some that it could be done very rapidly. If it has not happened to everyone's satisfaction in the past forty years, then even with a climate change in the engineering and system development disciplines it will still take time and resources to build adequate measures and methods.

There are some key criteria which need to be met if human factors is ever to become a full partner in certification of aviation systems. The first criteria to be met involves the pursuit of organizational change. Through a Total Quality Management emphasis or through an alternative educational process as suggested by Booher (1990), we need to produce organizational change and recognition that human factors is needed and is valuable. This may not be something you can mandate as the Army has tried but rather should involve attitude change out of self interest on the part of system developers. While human factors has come along way in the aviation community, there are still many who see it as a soft science, if they view it as a science at all.

Along with the organizational change regarding human factors, there needs to be some change in attitude in terms of accepting that there may be better ways of doing things in the development of new technology. The doctrine behind systems engineering is workable if the systems engineers remember to include the operators in their designs and if they seek out technical help related to person-machine issues. Human factors should be a system life cycle issue and not something that is only considered for the short run until the first prototypes are fielded. This type of thinking will show the rewards and costs of system use throughout their life expectancy and not merely the here and now. In many cases such costs are not considered; we live with systems that are user unfriendly and less efficient than they could be.

As human factors professionals, a critical criteria should be to never promise more than we can deliver and to deliver all that we promise. In order to do this, we have to be in the dual role of helping systems developers today avoid the obvious errors of design while constantly trying to develop new and better measurement tools, which are empirical and tailorable to specific applications. Measurement tools must be reliable and valid against systems goals. They also must meet the test of face validity if they are to ever be employed.

## References

- U. S. Army (1990). *AR 602-2 Manpower and personnel integration (MANPRINT)*. Washington, DC: Department of the Army.
- Boorer, H. R. (1990). *MANPRINT an approach to systems integration*. New York: Van Nostrand Reinhold.
- FAA (1991). *The national plan for human factors (draft)*. Washington, DC: Federal Aviation Administration.
- Fitts, P. M. & Jones, R. H. (1947). Analysis contributing to 460 "pilot error" experiences in operating aircraft controls. In H. W. Sinaiki (Ed.). *Selected papers on human factors in the use and design of control systems*. New York: Dover.
- Kantowitz, B. H. & Sorkin, R. D. (1983). *Human factors*. New York: Wiley.
- DOD (1981). *MIL-STD 1472D Military standard human engineering design criteria for military systems, equipment and facilities*. Washington, DC: Department of Defense.
- Nagel, D. C. (1988). Human error in aviation operations. In E. L. Weiner and D. C. Nagel (Eds.). *Human factors in aviation*. San Diego: Academic Press.
- Stein, E. S. (1985). *Air traffic controller workload: An evaluation of workload probe (DOT/FAA/CT-TN84/24)*. Atlantic City, NJ: DOT/FAA Technical Center.
- Thomas, D. D. (1985). ATC in transition, 1956-1963. *Journal of ATC*, 30-38.
- Warm, D. D. & Dember, W. N. (1986). Awake at the switch. *Psychology Today*, 20(4), 46-53.